DOCUMENT RESUME

ED 200 987 CS 206 258

AUTHOR Dilworth, Collett B., Jr.
TITLE Predilection in the Assessment of Writing.
PUB DATE Mar 81
NOTE 25p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (32nd, Dallas, TX, March 26-28, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College English; *Evaluation Criteria; Higher Education; Individual Characteristics; *Teacher Attitudes; *Writing Evaluation; Writing Instruction; *Writing Research
IDENTIFIERS English Teachers

ABSTRACT

A study was conducted to examine how composition characteristics identified on the Diederich scale (ideas, organization, wording, flavor, usage/sentence structure, and punctuation) related to teacher evaluation of student writing. For each of three academic years, teaching assistants of freshman English were given typed, mimeographed copies of student essays to rate using a modified Diederich scale (without the spelling and handwriting sections). Over the three-year period, 60 teachers rated 13 essays, and the ratings were analyzed for patterns of preference among the teachers to identify high (liberal) and low (conservative) raters for each of the six items of the Diederich scale. Over three years, the only consistent, statistically significant pattern was for teachers who tended to rate "ideas" conservatively (preferring logical, analytical, relatively abstract essays) or liberally (preferring impressionistic, imagistic essays). In the second part of the study, two of the differentiating essays were rated by 14 full-time English department faculty members. The results supported the idea that evaluators are significantly influenced by predilections for either ratiocinative or impressionistic prose. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED200987

PREDILECTION IN THE ASSESSMENT OF WRITING

Collett B. Dilworth, Jr.
Department of English
East Carolina University
Greenville, North Carolina 27834
Phone: 919-757-6041

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Collett B. Dilworth

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS206258

PREDILECTION IN THE
ASSESSMENT OF WRITING

Ever since Diederich, French, and Carlton's factor analysis (1961), the problem of how composition characteristics relate to teacher evaluation has been investigated primarily in terms of the variables identified in that study and subsquently refined by Diederich (1974): ideas, organization, wording, flavor, usage/ sentence structure, punctuation, spelling, and manuscript form. Diederich and his colleagues derived their factors from an analysis of the reasons a sample of evaluators gave for the grades they assigned to a collection of student essays. The analysis suggested that the factors identify certain "schools of thought" or predilections which variously influence evaluator judgment.

Of the factors, ideas and mechanics (usage, sentence structure, spelling, punctuation) accounted for most of the variance in ratings. Evidence for a predominant influence of ideas and organization is reported by Freedman (1979) and Harris (1977). Nold and Freedman (1977) found that the influence which sentence structure does exert may in large be accounted for by incidence of free modifiers, length, and (negatively) by incidence of auxiliary "be" and "have." Neilsen and Piché (1981), however, found that sentence structure was not a predictor of quality rating but that wording was. Investigating possible reasons for rater predilection, Eldridge (1981) offers longitudinal evidence that changing social conditions and values affect evaluators' perspectives. By comparing the ratings and comments that seven college faculty colleagues gave to an essay in 1972 with the

ratings and comments they gave to the same essay in 1978, Eldridge suggests that the profession has indeed experienced a "shift in focus from content-oriented to form oriented evaluation."

In all such studies, evaluator biases are identified in terms of contraries: that is, qualities defining the limits of good vs. bad in particular rhetorical dimensions. For example, Freedman (1979) defines poor "content" as that containing misinterpretation of topic, irrelevance of ideas, dysfunctional repetition, illogical contradictions and ambiguities. She implicitly defines good content in terms contrary to these.

DeBeaugrande (1979), on the other hand, has made an effort to keep general quality constant in designing six versions of an essay on a rocket's flight. Rather than contrary qualities, these essays manifest differences in kind. The six versions were rated and commented on by a group of experienced teachers. Two of the six essays elicited a sharp pattern of preference. Eight teachers preferred the first because of its vividness, and criticised the second because of its lack of color and flavor; nine teachers conversely favored the second because of its efficiency, and criticised the first because of its elaborate imagery. DeBeaugrande consequently identifies two "ideals": conciseness and vividness. While they are pertinent to Diederich's category of "flavor," these ideals were not differentiated by a standard rating of good or bad flavor. Instead the raters seemed to be expressing a preference for one kind of flavor over another.

Procedures

The present study further investigates how evaluator

predilection is manifest as preference for difference in kind rather than as discernment of difference in degree. During each academic year for three years, teaching assistants of freshman English at a state university were given typed, mimeographed copies of student essays to rate using a modified Diederich scale, the spelling and handwriting catagories ommitted (Diederich, 1974). In each case the TA's had been trained in the use of the scale for approximately one and a half hours before they read the essays. In the first year 27 TA's rated six essays, in the second year 16 TA's read three essays, and in the third year 17 TA's read four essays. Besides rating analytically, the teachers also ranked the essays according to a holistic evaluation of quality.

The 13 essays were selected by the researcher with the help of faculty colleagues according to these criteria:

1. They should be 300-400 word first semester freshman English essays, planned out of class, written in one class hour and revised in a second class hour.

2. They should have an ambiguous overall quality with the potential to elicit a range in ratings.

3. In the second year all essays were "definitions," and in the third year all were "comparison/contrasts."

Analysis of the data

Each year the ratings and rankings were analysed to discern patterns of preference among the teachers. The strategy was to identify high (liberal) and low (conservative) raters for each of the six items on the modified Diederich scale and for the scale as a whole. High raters on any given scale item were defined as

those whose total ratings for the essays were in the highest third of the totals, and low raters were defined as those whose total ratings for the essays were in the lowest third. For example, to select the liberal and conservative raters of "ideas" in the first year, the total ratings of ideas for the six essays was computed for each of the 27 teachers and the highest nine and lowest nine were identified.

For the high and low raters on each scale item, the mean rank given each essay was also computed. The ranks of these means for high and low raters were then compared for differences. When two mean rankings for an essay differed by 1.0 or more the difference in preference by the two groups was tested by chi square. Chi square was used because the data is ordinal. But because the N's are less than the N's conventionally recommended for non-parametric analysis, statistical significance was used as an arbitrary threshold for the admission of evidence. Rejecting an hypothesis of no difference in this study, therefore, depends as much on the consistant replication of difference as on the probability of sampling error. To illustrate the procedure for initial identification of difference, Table 1 shows a comparison of mean rankings for the six essays read in the first year by high and low raters on ideas.

Put Table 1 about here.

This table shows that high raters on ideas tended to rank the essay "My Favorite Place" over "Pinehurst and Aberdeen," and that low raters tended to rank "Pinehurst and Aberdeen" over "My Favorite

Place." (The crossbreak and test for significance of this particular pattern are shown in the left portion of Table 2, below.) Table 1 also shows that the low raters strongly tended to rank the essay "Backgammon" more highly than did the high raters.

Results

Analyzing the ratings in this way revealed only one regular pattern, and this pattern was statistically significant for the first and third years, just missing significance in the second year: Consistantly over the three years, high and low raters on ideas tended to differ in their rankings of essays. None of the other high-low comparisons for organization, wording, flavor, usage/sentence structure, punctuation, or total scale score was consistant in pattern of preference over the three rating sessions, nor were any of their differences in preference statistically significant.

Table 2 shows the chi square analysis of the two pair of significantly different relative rankings by high and low raters on ideas.

Put Table 2 about here.

This table shows that at statistically significant levels the high and low raters on ideas tend as groups to disagree on the relative merit of specific pairs of essays.

In addition to the reversed rankings for a pair of essays, the raters on ideas in the first session differed significantly in their rating of one other essay, "Backgammon." This difference is shown in Table 3.

Put Table 3 about here.

In this table we see the marked tendency for conservative raters on ideas to rank the essay on backgammon more highly than do the liberal raters.

The Discriminating Essays

The fact that no such significant differences were found in the preferences by groups of high and low raters on the five other scale items suggests that relatively liberal and relatively conservative approaches to evaluating ideas are singularly related to rater predilection. To discern what other variables seem to be involved, this inquiry compared the essays which had elicited the different rating patterns. Below are the crucial paragraphs from the two reverse ranked essays of the first year.

From "Why Pinehurst and Aberdeen Are Diverse in Character"

> One of the major features that determines the image of any single town is the availability of spending money among the populace. We can see this by observing the physical appearance of the municipality. First, let us look at Aberdeen. A short drive through the residential districts will reveal that the incomes, as reflected in the housing, range from low middle class to middle class. Driving through the business district, we see the items for sale, the condition of the shops, and even the salespersons themselves reinforce the image of a middle class township

with a little "country hick" thrown in for good measure. Now if you drive through Pinehurst, the picture we encounter is quite different. The housing is that of people with substiantially more than a middle class or even upper middle class income. In many cases, the homes are virtually mansions. If we visit one of the many emporiums, we will find that the goods and their corresponding prices, the atmosphere, and salespeople all cater to the desires of the wealthy....

From "My Favorite Place"

Ocracoke is a very small island and has a population of only three hundred. The island itself is thirteen miles long and two miles wide (in the widest spot). The village only covers a very small part of the island as most of the island is marshes, beaches, or other underdeveloped areas. In the village you will find one crowded general store, three fishy restaurants, four hotels, one sleepy gas station, one policeman, and a one-engine fire department. With the exception of the general store, none of these things existed twenty years ago. Some things that still don't exist are stop lights, taxis, city water, and smoke stacks. Now you might think this is a pretty dull place. Actually it's not; the lack of these things make it all the more unique....

Both of these essays were highly rated by teachers from both groups. They are similar, of course, in that they treat the specifics of communities, but their differences are suggestive.

"Why Pinehurst and Aberdeen Are Diverse in Character" (PA), preferred by conservative raters on ideas, is a purposeful comparison of two towns, whereas "My Favorite Place" (FP), preferred by liberal raters on ideas, is a description of a single community. Although both essays mention concrete aspects of locale, FP is more specific and imagistic; PA has no equivalents for FP's fishy restaurants, sleepy gas station, and one-engine fire department. On the other hand, PA develops a line of logically related evidence; PA's writer begins with a claim and proceeds to try systematically to establish it. FP's writer, in contrast, offers details not as supportive evidence, but as sensory impressions for their own sake. Indeed he neglects to demonstrate the logic of his claim proposed in the last two sentences quoted, for nowhere does he show how Ocracoke's "uniqueness" should not make it "dull." For this writer the claim seems to be an afterthought, a sop to the traditional requirement for a thesis. That these two essays elicited different rankings suggests that the two groups of raters may differ in their predilections for imagistic, mosaic prose and for logical, analytic prose.

Scruitiny of the other essay which elicited a difference in rankings by high and low raters on ideas lends support for this hypothesis. An excerpt is given below.

From "Backgammon"

> To begin the game each player rolls one die and the player with the highest number moves the combination of the two dice. Only if the dice were the same do you roll again for the beginning move. After

> the first roll, players roll both of their dice and move their alloted moves as shown on the dice, but when two or more of the opponent's men are on a point the player cannot land on that point. Keeping this in mind, players try to move with either two men or land safely on one or more of their own.

This paper is a complex but reasonably well ordered discussion of how to play backgammon. Such a discussion requires the reader to think in terms of sequences, to comprehend "if-then" relationships, and to note exceptions and alternatives. Of the six essays rated the first year, this one is the least impressionistic and the most logically linear. That it was rated realtively high by the conservative raters on ideas and relatively low by the liberal raters could be predicted in light of the preference pattern discussed above.

The following year, restrictions on time precluded wide reading of essays in meetings with new TA's, still the preference patterns for the three rated essays replicated the earlier patterns. Thus by the third year an hypothesis that rating of ideas relates to predilection seemed strongly tenable, and so of the four essays read by the TA's that year, two were selected which thoroughly manifested the impressionistic and the rational. Excerpts from the essays are quoted below.

From "Summer Job Prospects"

> This summer I have the choice between two different employment options, one which offers guarantees and one which offers possibilities. This decision may change the shape of my future.

The first job is employment in a restaurant. It offers me the financial security of minimum wage and the average of forty working hours each week. These hours still allow me plenty of spare time to enjoy summer.... Chosing this job means living at my parents' house, enjoying family and not worrying about spending money for rent or meals.

The second job is employment as a door-to-door book salesperson. It offers me eighty working hours each week and a wage based on commissions, which only means a fluctuating income with the possibility of good money.... This is an out of state job, which means ... I will be my own boss, paying my own bills, solving my own problems, and living in society as a true adult.

After evaluating the situation, I have reached a decision. I'm not making the choice on the chance of making lots of money, but on the chance that I would be gaining valuable knowledge by tasting the independence and responsibility that can only be obtained by sacrifice of time. I choose the salesperson job....

From "Guitars"

There are two main types of electric guitars, the solid-body and the hollow-body. Though the two may seem similar in appearance, each has a different and distinct personality.

The solid-body electric is a guitar carved out of

a single slab of wood. A hard wood like maple is used because the density of the body gives a guitar the ability to sustain notes for an indefinite length of time....

The solid-body has the sleek, angular, flashy look that also characterizes the music that it was meant to play. Just one sweep of the strings and you're overcome with the sense of raw, untamed power....

The hollow-body guitar is like a hollow egg, the shell constructed of a thin layer of rosewood and nothing but space within its walls. The air vibrates producing a soft, smooth sound....

The hollow-body has a smooth, curved appearance, breathing easy-flowing melodies and weeping bitter sweet notes that melt into your memory.

The rough and ready-for-action fighter and the sentimental-visioned dreamer are the most beautiful creations in this universe when in the hands of a person that really understands and loves.

Both of these essays compare and contrast phenomena important to the writers, but "Summer Job Prospects" (SJP) develops a line of reasoning while "Guitars" (G) accrues imagery less to support a thesis than to demonstrate impressions. The logic of SJP is explicit, its persona openly evolving a rationale. In G the writer's impressions do relate to a generalization, that each type of guitar "has a different and distinct personality,"

but if we omit the generalization we note minimal effect on the essay. In fact, deleting any sentence from the last five paragraphs of G has the effect not of interrupting something but of lessening imagistic detail and vividness; whereas if we delete any sentence from SJP we do not weaken imagery, but we do interrupt a chain of evidence. The two cells on the right of Table 2 show a clear preference for SJP by low raters on ideas and for G by high raters. This pattern is consistant with the earlier pattern.

Interviews with Raters

The Diederich scale responses, however, have limited worth as independent variables, for characterizing a rater as liberal or conservative does not portray the individual to any depth. One obvious comprehensive explanation for the responses is offered by the research in brain hemisphericity (Ornstein, 1972). Viewed from this perspective, the importance of ratiocination to the raters correlates with how liberally or conservatively they evaluate ideas: the greater the influence of the left hemisphere, the more rigorous the evaluation on the basis of the quality of reasoning, and vice versa. Apparently the reasoning in PA, SJP, and "Backgammon" appeals to the left brain, and the impressionism in FP and G appeals to the right brain.

To explore further the predilections suggested by the grouped data and to identify specific terms of logical and impressionistic response, the investigator elicited composition evaluations individually from fourteen volunteering full time members of the English faculty at his university. The essays evaluated were "Summer Job Prospects" and "Guitars." The procedure was for each rater to read the essays, to rank them, to grade them, and to

discuss with the investigator the reasons for preferring one essay to the other. During the interviews, the investigator explained the procedure but did not guide the disscussion in any way other than to request in general terms more information. These requests were to the effect, "Can you tell me more about your preference? Do you have anything else to say about this essay?"

Of the fourteen faculty, six preferred SJP and eight preferred G. As a group the six preferring SJP rated it only slightly higher than they rated G (B+ average vs. B average). Those who preferred G, however, gave that essay a much higher relative average grade (B+ vs. C). None of those preferring SJP were explicitly enthusiastic about the essays, but five of those preferring G used intensifiers in such contexts as, "I <u>really</u> liked that" and "This part is <u>quite</u> good." It may be that the six preferring SJP, while predisposed to analytic prose, found G to be so strong an example of impressionistic prose that they differentiated relatively little in merit between the two. In fact, ten of the fourteen faculty members suggested that the author of G had more potential as a writer than did the author of SJP.

On the other hand, all six of the SJP group, although recognizing the vitality of G, complained of excessive impressionism:

SJP #1: "a lyric quality here; perhaps too lyric"

SJP #2: "the sharp, facile, strained picturesqueness"

SJP #3: "purple prose"

SJP #4: "too reliant on the writer's responses; too egocentric"

SJP #5: "excessively impressionistic"

SJP #6: "overstrained emotionally"

For the SJP group, the organization and the logic of their preferred essay are its most attractive features. Five of the six considered the paragraphing in SJP to be much superior to that of G, and the same number particularly liked tha sequential reasoning of SJP:

SJP #1: "the careful reasoning"

SJP #2: "the specific facts leading to a conclusion"

SJP #3: "the reasoning"

SJP #5: "the thinking that goes on"

SJP #6: "the problem solving process"

One who preferred SJP, however, did not like the "gimmicky nature of the decision process"(SJP # 4).

The eight faculty who preferred G were unanimous in their praise of the concreteness, the imagery, and the "vitality" of that essay:

G #1: "use of figurative language and specifics"

G #2: "clear, picturesque terms describing function and appearance"

G #3: "tangible detail"

G #4: "specificity"

G #5: "convergence of sense impressions"

G #6: "specificity and concreteness"

G #7: "metaphors"

G #8: "figurative devices"

Six of the eight were explicit in preferring the diction of G to that of SJP, but none of the eight expressed concern with the paragraphing in the essays, unlike the faculty preferring SJP.

The faculty preferring G were most negative about what they perceived to be the abstractness and vacuity of SJP.

G #1: "not the same attention to creating something interesting for the reader"

G #2: "The internal struggle over the decision is unconvincing."

G #3: "not as attractive; it repels me rather than draws me to it."

G #4: "less sophisticated, more abstract"

G #5: " superficial; no meat"

G #6: "a complicated task not carried off"

G #7: "less specific; more pedestrian"

G #8: "dull; no real communication in the sense of an effort to draw an audience to the topic"

The interviews confirm that the appeal of SJP rests in its analytic approach and in its linking of facts and abstractions ("eighty working hours", "possibility", "living in society as a true adult"), and that the appeal of G rests in its concrete, suggestive approach and its evocation of sensory impressions ("sleek, angular, flashy"). These are among the appeals that distinguish the responses of the two brain hemispheres.

Conclusion

The preferences for order, logic, and explicit sequence by certain raters and for imagery and cumulative impressions by certain other raters lend support to the hypothesis of there being at least two major types of predilection among composition teachers. Indeed, the consistancy and ease of replicating

De Beaugrande's two "ideals" suggest that they are fundamental. But of course their influence is not indomitable. Meyers, McConville, and Coffman (1966) and many other studies (as well as extensive clinical experience) demonstrate that the effects of individual reader bias can be neutralized by structuring the writing assignment and by organizing and training evaluators. In such cases evaluators apply criteria which either standardize or integrate the influence of specific biases. For example, primary trait scoring (Lloyd-Jones, 1977) is an especially thorough attempt to eliminate altogether the effects of singular bias. The tightly channeled assignments and thoroughly delineated levels of achievement pertinent to the assignments insure high inter-rater reliability for primary trait scorers. For all such efforts, though, heightened reliability is purchased at the expense of student and teacher initiative, because the more the discretion left to individual students and teachers, the more singular their interaction and the less replicable the specific evaluations of student writing. Furthermore, predilections for the logical and for the impressionistic on the part of writers probably affect the quality of their responses to highly structured assignments. And since primary trait and analytic scale scoring in effect establish official predilections, writing that appeals to diverging predilections will inevitably be penalized.

In sum, the results of the present study suggest that in normal writing curricula, logical and impressionistic perspectives do influence teacher response to student work. Very probably these values held by students and teachers interact in particular

ways. We can conceive that teachers especially attracted either to ratiocination or to impressionism might structure assignments which particularly encourage their preferred type of prose and might especially nurture students who write it. When opposite types interact, the results might vary from a broadening of students' abilities to a general discouraging of their efforts. In any case, we would do well to inform our writing instruction and our research with a greater awareness of our predilections.

References

DeBeaugrande, R. Psychology and composition. College Composition and Communication, 1979, 30, 50-57.

Diederich, D. B. Measuring growth in english. Urbana, Illinois: National Council of Teachers of English, 1974.

Diederich, P.B., French, S., & Carlton, S. Factors in judgements of writing ability (research bulletin 61-15). Princeton, NJ: Educational Testing Service, 1961.

Eldridge, R. Grading in the 70's: how we changed. College English, 1981, 43, 64-68.

Freedman, S. W. How characteristics of student essays influence teachers' evaluations. Journal of Educational Psychology, 1979, 71, 328-338.

Harris, W. H. Teacher response to student writing: a study of the response patterns of high school english teachers to determine the basis for teacher judgement of student writing. Research in the Teaching of English, 1977, 11, 175-185.

Lloyd-Jones, R. Primary trait scoring, in Evaluating writing, C. Cooper & L. Odell, eds. Urbana, Illinois: National Council of Teachers of English, 1977, 33-66.

Meyers, A., McConville, C. & Coffman, W. Simplex structure in the grading of essay tests. Educational and Psychological Measurement, 1966, 26, 41-54.

Neilsen, L. & Piché, G. L. The influence of headed nominal complexity and lexical choice on teachers' evaluations of writing. Research in the Teaching of English, 1981, 15, 65-73.

Nold, E. & Freedman, S. W. An analysis of readers' responses to essays. Research in the Teaching of English, 1977, 11, 164-174.

Ornstein, R. E. The psychology of consciousness. San Francisco: Freemsn, 1972.

TABLE 1

Mean Rankings of Six Essays by
High and Low Raters on Ideas

Essay	"My Favorite Place"	"Pinehurst and Aberdeen"	"Writing and Painting"	"Backgammon"	"Making a Dress"	"Taming the Beast"
High Raters on Ideas	1.66	2.50	2.57	4.50	4.67	5.17
Low Raters on Ideas	2.83	1.50	3.14	3.16	4.50	5.83

TABLE 2

Preference for
Two Pairs of Essays by
High and Low Raters on Ideas

	First Rating Session		Third Rating Session	
Essay Preferred	"A Favorite Place"	"Pinehurst and Aberdeen"	"Summer Job Prospects"	"Two Guitars"
High Raters on Ideas	8	1	5	1
Low Raters on Ideas	2	7	1	5
	$X^2 = 8.10$, $p < .01$		$X^2 = 5.33$, $p < .05$	

TABLE 3

Ranking of "Backgammon" by High Raters on Ideas and Low Raters on Ideas

	"Backgammon" Ranked in Top Half	"Backgammon" Ranked in Bottom Half
High Raters on Ideas	1	8
Low Raters on Ideas	5	4

$X^2=4.00$, $p<.05$